

COSMOGRAIL: the COSmological MONitoring of GRAVItational Lenses^{★,★★}

XII. Time delays and 9-yr optical monitoring of the lensed quasar RX J1131–1231

M. Tewes¹, F. Courbin¹, G. Meylan¹, C. S. Kochanek^{2,3}, E. Eulaers⁴, N. Cantale¹, A. M. Mosquera^{2,3}, P. Magain⁴, H. Van Winckel⁵, D. Sluse⁶, G. Cataldi¹, D. Vörös¹, and S. Dye⁷

¹ Laboratoire d'astrophysique, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland, e-mail: malte.tewes@epfl.ch

² Department of Astronomy, The Ohio State University, 140 West 18th Av., Columbus, OH 43210, USA

³ Center for Cosmology and Astroparticle Physics, The Ohio State University, 191 West Woodruff Av., Columbus, OH 43210, USA

⁴ Institut d'Astrophysique et de Géophysique, Université de Liège, Allée du 6 Août, 17, 4000 Sart Tilman, Liège 1, Belgium

⁵ Instituut voor Sterrenkunde, Katholieke Universiteit Leuven, Celestijnenlaan 200B, 3001 Heverlee, Belgium

⁶ Argelander-Institut für Astronomie, Auf dem Hügel 71, 53121, Bonn, Germany

⁷ School of Physics and Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD, UK

Received ; accepted

ABSTRACT

We present the results from 9 years of optically monitoring the gravitationally lensed $z_{\text{QSO}} = 0.658$ quasar RX J1131–1231. The R band light curves of the 4 individual images of the quasar are obtained using deconvolution photometry, for a total of 707 epochs. Several sharp quasar variability features strongly constrain the time delays between the quasar images. Using three different numerical techniques, we measure these delays for all possible pairs of quasar images, while always processing the 4 light curves simultaneously. For all three methods, the delays between the 3 close images A, B and C are compatible with being 0, while we measure the delay of image D to be 91 days, with a fractional uncertainty of 1.5% (1σ), including systematic errors. Our analysis of random and systematic errors accounts in a realistic way for the observed quasar variability, fluctuating microlensing magnification over a broad range of temporal scales, noise properties, and seasonal gaps. Finally, we find that our time delay measurement methods yield compatible results when applied to subsets of the data.

Key words. Gravitational lensing: strong – cosmological parameters

1. Introduction

Using the time delays between strong lensing images to measure cosmological distances (Refsdal 1964) has a number of advantages: there is no need for any primary or secondary calibrator, and there are no effects from the intergalactic or interstellar medium. The method, originally proposed for gravitationally lensed supernovae, has so far exclusively been applied to quasars lensed in most cases by individual massive galaxies. Exceptions are SDSS J1004+4112 and J1029+2623, two quasars lensed by galaxy clusters, with long time delays (Fohlmeister et al. 2008, 2012). The quasar lens time delay method is now recognized as a tool that complements other cosmological probes, in particular to constrain H_0 as well as the dark energy equation of state parameter, w (e.g., Suyu et al. 2012b; Linder 2011; Moustakas et al. 2009). In spite of its advantages, the method has long faced two severe limitations to its effectiveness in constraining cosmology.

First, time delays between the gravitationally lensed images of a quasar are hard to measure. The first time delay measurements were quite controversial (Vanderriest et al. 1989; Press et al. 1992; Schild & Thomson 1997; Kundic et al. 1997). Understandably, early light curves tended to be short and sparse, often too short to clearly demonstrate that *microlensing* variability was not interfering with their analysis. Microlensing is seen as an uncorrelated extrinsic variability in the quasar images, which results from the time-variable magnification created by stars in the lensing galaxy (e.g., Chang & Refsdal 1979; Schmidt & Wambsganss 2010). In the best cases, light curves spanned a few years (see, e.g., Wyrzykowski et al. 2003; Hjorth et al. 2002; Burud et al. 2002a,b). One consequence is that the numerical methods used to measure time delays from these light curves were exceedingly “optimistic” in their assumptions about extrinsic variability. Later measurements with better data frequently yielded delays inconsistent with the error estimates of the earlier measurements.

Second, given measured time delays, and lens and quasar image astrometry, there is a famed degeneracy between the *time delay distance*, which is a scale parameter inversely proportional to H_0 , and the spatial distribution of mass responsible for the strong lensing phenomenon. The delays constrain only a combination of this time delay distance and the surface density of the lens near the images (Kochanek 2002). This can be over-

[★] Based on observations made with the 1.2-m Swiss Euler telescope (La Silla, Chile), the 1.3-m SMARTS telescope (Las Campanas, Chile), and the 1.2-m Mercator Telescope. Mercator is operated on the island of La Palma by the Flemish Community, at the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofísica de Canarias.

^{★★} Light curves will be available at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.128.5) or via <http://cdsarc.u-strasbg.fr/viz-bin/qcat?J/A+A/???>, and on <http://www.cosmograil.org>.

come given independent constraints on the structure of the lens. Suyu et al. (2009, 2010) convincingly show that it is possible to control the effects of model degeneracies in the case of B1608+656, a quadruply imaged quasar with accurate radio time delays (Fassnacht et al. 2002). To do this, the authors combine (1) detailed HST images of the lensed quasar host galaxy, (2) a velocity-dispersion measurement of the lens galaxy, and (3) information about the contribution of intervening galaxies along the line of sight, from galaxy number counts calibrated with numerical simulations.

In parallel to the advances in lens modeling, the observational situation has drastically evolved as well. Two observational groups, COSMOGRAIL and Kochanek et al. (2006), have been intensely monitoring ≈ 20 lenses for roughly 10 years. In 2010, our 2 groups decided to merge their observational efforts, with the COSMOGRAIL group focusing on the analysis of time delays and the Kochanek et al. (2006) group focusing on the analysis of microlensing. While preliminary results have been published both before and after this merger (Kochanek et al. 2006; Vuissoz et al. 2007, 2008; Morgan et al. 2008a,b), exquisite data spanning almost a decade of continuous observation are now being released, such as for the quadruply imaged quasar HE 0435-1223 (Courbin et al. 2011; Blackburne et al. 2011).

In this paper, we present 9 years of optical monitoring of the quadruply imaged quasar RX J1131–1231 (Sluse et al. 2003), and measure its time delays with the techniques of Tewes et al. (2012). RX J1131–1231 is one of the most spectacular lenses of our sample. The redshift of the lensing galaxy is $z_{\text{lens}} = 0.295$, while the quasar is at $z_{\text{QSO}} = 0.658$. This low quasar redshift means (1) that the photometric variations are fast, numerous and *strong* because it is a lower luminosity quasar (see, e.g., MacLeod et al. 2010), and (2) that the host galaxy of the lensed quasar is seen as a full Einstein ring with many spatially resolved structures in HST images. Similarly, the lensing galaxy is sufficiently bright to allow a precise measurement of its velocity dispersion and possibly of its velocity dispersion profile. These characteristics ease both the time delay measurement and the lens modeling. The latter, with state-of-the-art inferences of cosmological constraints based on our time delay measurements of RX J1131–1231, are presented in Suyu et al. (2012a).

The COSMOGRAIL and SMARTS observations of RX J1131–1231 and their reduction are described in Sections 2 & 3, while the light curves are presented in Section 4. In Section 5, we apply three different curve shifting techniques to the light curves and we infer our best measurements of the delays along with realistic random and systematic error bars. Our results are summarized in Section 6.

2. Observations

We have been monitoring the quadruply lensed quasar RX J1131–1231 (J2000: $11^{\text{h}}31^{\text{m}}52^{\text{s}}$, $-12^{\circ}31'59''$) since December 2003 with 3 different telescopes in the R band ($\sim 600 - 720$ nm). Table 1 summarizes the observational strategy and instrumental characteristics. The light curves presented in this paper cover 9 observing seasons (2004 - 2012, see Fig. 4), with 707 monitoring epochs in total. The average sampling within the seasons is 2.9 days, and the median sampling is 2.0 days. The mean seasonal gap in the combined light curves is 132 days with a standard deviation of 2 weeks.

The majority of the measurements for this southern target came from the Swiss 1.2-m Euler telescope and the SMARTS

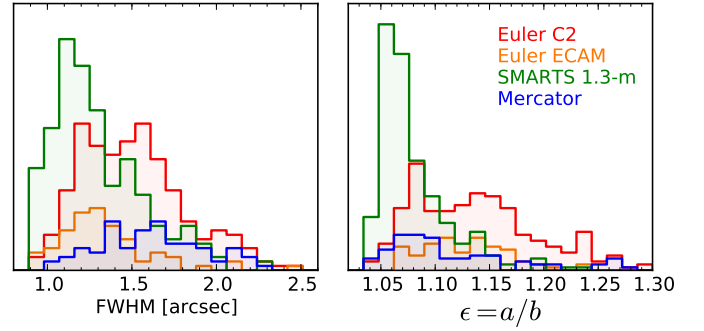


Fig. 1. Distributions of stellar FWHM and elongation ϵ of all images contributing to the light curves, by telescope.

1.3-m telescope, both located in Chile under equally good atmospheric conditions. Fig. 1 shows the distributions of the stellar Full Width at Half Maximum (FWHM) and the elongation ϵ of each observing epoch, as measured by SExtractor (Bertin & Arnouts 2010) using field stars. The SMARTS 1.3-m telescope is guided, in contrast to Euler and Mercator, which are solely tracking. This accounts in part for the broad elongation distribution of the Euler and Mercator images.

The original imaging instrument C2 of the Euler telescope was replaced in September 2010 by *EulerCAM*, a liquid nitrogen cooled $4\text{k} \times 4\text{k}$ e2v 231-84 CCD yielding a pixel scale of $0''.215$. At the same time the focusing procedure was improved. Fig. 1 shows that after two years of observations, images from *EulerCAM* tend to be statistically sharper than C2 images, and more comparable to SMARTS 1.3-m data in terms of FWHM. All images from the SMARTS 1.3-m telescope were obtained through the optical channel of the ANDICAM¹ camera (DePoy et al. 2003). See Kochanek et al. (2006) for details about the SMARTS data.

Fig. 2 shows the central part of a deep combination of Euler monitoring images, totaling 2.5 days of exposure. In the best individual exposures the 4 quasar images are well resolved; images A and B have the smallest separation of $1''.2$.

3. Image reduction

In this section we describe the reduction procedure leading to the light curves of the multiple quasar images of RX J1131–1231. This same procedure will homogeneously be applied to other lenses monitored by COSMOGRAIL and SMARTS, hence we provide a general description of our methods.

All exposures are corrected for bias/readout effects and flat fielded using standard methods. This prereduction is done individually for each telescope and instrument. For Euler and Mercator we have developed custom python pipelines that enable efficient human inspection and validation of each step. Particular attention is given to a semi-automated selection and combination of sky-flats, which is controlled through a simple graphical interface highlighting the temporal evolution of instrumental contaminations such as dust. Using difference images we check that all sources in the flatfield exposures are adequately masked and that the contamination does not evolve significantly within sets of flatfields that are stacked as master-flats. The period over which we stack flatfields can be from single days to several weeks. Some exposures of the Mercator telescope are prereduced using “superflats”, combinations of masked science

¹ <http://www.astronomy.ohio-state.edu/ANDICAM/>

Table 1. Overview of our optical monitoring of RX J1131–1231

Telescope	Location	Instrument	Pixel scale	Field	Exp. time	Epochs	Sampling
Euler 1.2 m	ESO La Silla, Chile	C2	0''344	11' × 11'	5 × 360s	265	5.0 (4.0)
Euler 1.2 m	ESO La Silla, Chile	EulerCAM	0''215	14' × 14'	5 × 360s	76	5.4 (4.1)
SMARTS 1.3 m	CTIO, Chile	ANDICAM	0''369	6.1' × 6.1'	3 × 300s	288	6.1 (5.1)
Mercator 1.2 m	La Palma, Canary Islands, Spain	MEROPE	0''193	6.5' × 6.5'	5 × 360s	78	8.9 (4.5)

Notes. The column “Exp. time” indicates the number of dithered exposures *per epoch* and their individual exposure times. One epoch corresponds to one data point in each light curve. The temporal sampling of the observations is given as the average (median) number of days between consecutive epochs, excluding the seasonal gaps.

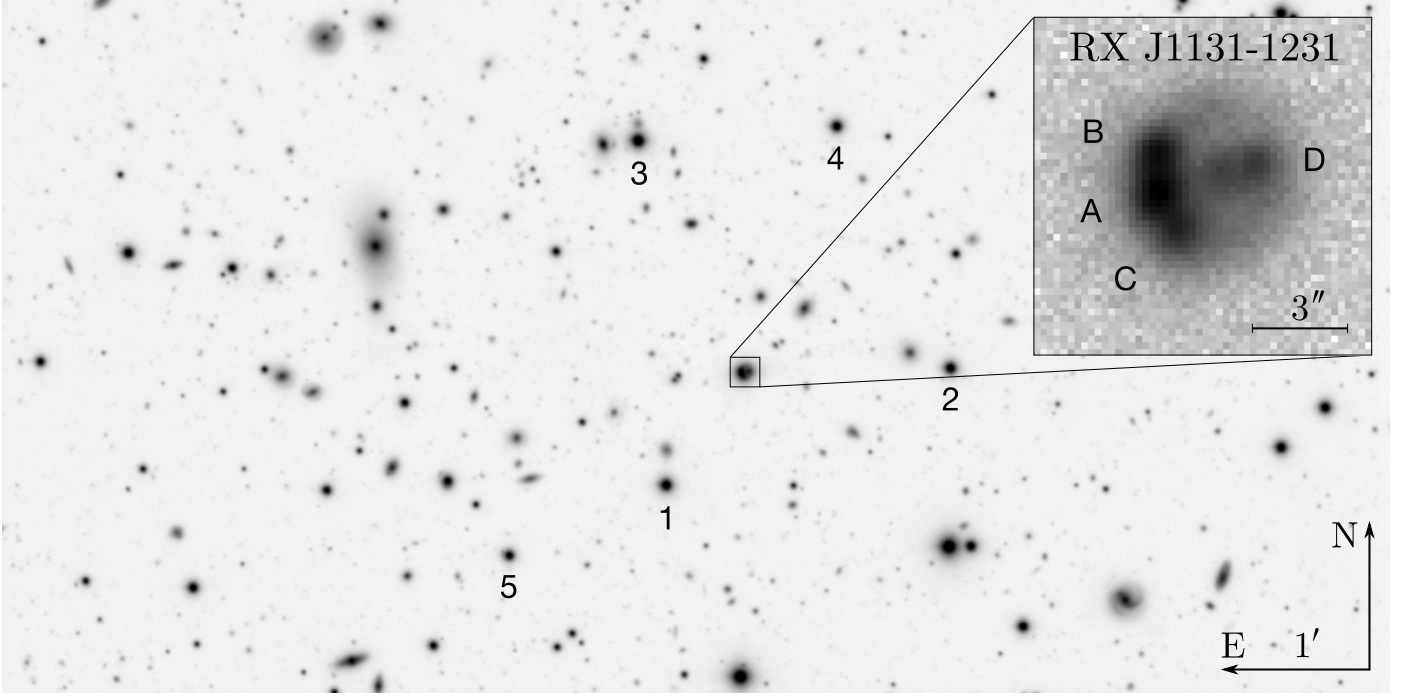


Fig. 2. Part of the field of view of the Swiss 1.2-m Euler telescope around RX J1131–1231. The wide-field image is a combination of 600 exposures of 360s each, corresponding to a total exposure time of 2.5 days. The inset shows a single 360s exposure from the *EulerCAM* instrument under good conditions (FWHM~1'', $\epsilon = 0.1$, no Moon). All images are taken in the R band. We mainly use the stars labeled 1 to 4 to build a PSF model for each exposure. Stars 1 to 5 are used for the photometric calibration.

exposures obtained in a single night. We have developed variants of these pipelines adapted to the particularities of other COSMOGRAIL telescopes.

From here on, all images are processed by a single deconvolution photometry software package. The principle ideas for this reduction procedure are the same as employed for previously published COSMOGRAIL light curves (Vuissoz et al. 2007, 2008; Courbin et al. 2011). Over the years we have reworked the steps, and implemented the procedure in the form of a *python* pipeline linked to a relational database, containing one entry per exposure.

3.1. PSF construction

We first build a conservatively smooth sky model, obtained using SExtractor, for each exposure. This sky model is not critical, as the subsequent photometry procedure will fit for any residual sky level around all sources of interest.

We then align the images, separately for each instrument, and individually estimate the Point Spread Function (PSF) of each exposure. This is done by fitting several field stars using a

common model, composed of (i) a simply parametrized profile and (ii) a regularized fine pixel array. The details of this procedure, which is part of a general purpose deconvolution package, are described in Cantale et al. (in prep). For most of the exposures of RX J1131–1231, we use the stars labeled 1 - 4 in Fig. 2 to build the PSF model. The pipeline allows us to easily explore and compare different choices of PSF-stars. The final choice of stars is empirically selected so as to yield the least scatter in the final quasar light curves. In the case of RX J1131–1231 the situation is close to optimal, since (1) the stars 1 - 4 are bright but still in the linear regime of the CCDs, (2) they surround the lens at modest angular distances, and (3) all stellar companions or background objects can be identified and masked, yielding clean fitting residuals. Cosmic ray hits and CCD artifacts are automatically masked using a variant of the L.A. Cosmic algorithm (van Dokkum 2001). We visually supervise the PSF construction and manually adapt the selection of PSF-stars for problematic frames. The construction of the PSF models is the predominant technical issue affecting the final quality of the light curves.

3.2. Photometric normalization of the exposures

We next compute a multiplicative normalization coefficient for each exposure, based on photometry of field stars. These coefficients will constitute the reference for the differential photometry of the quasar images.

In a first step, this computation is done separately for each instrument. We measure the instrumental fluxes $N_{\star ij}$ (in photons) of star i in each exposure j , by fitting the PSF $_j$ of this exposure to each star, leaving only the fluxes, astrometric shifts, and background levels of each star as free parameters. For each star and exposure, we then compute an individual calibration coefficient

$$c'_{ij} = \frac{\text{med}_j(N_{\star ij})}{N_{\star ij}}, \quad (1)$$

where med_j denotes the median over all exposures of this instrument. Then, for each exposure j , we obtain the normalization coefficient through

$$c_j = \text{med}_i(c'_{ij}). \quad (2)$$

The advantages of this simple and fast method is that it does not select a single exposure as the “reference” for all stars. For each instrument, the distribution of these coefficients c_j is highly unimodal, as the exposure time is kept constant.

To select the normalization stars we first inspect the normalized stellar light curves for any anomalies or variability on scales of months or years, and, if required, we iteratively repeat the coefficient computation with a refined selection of stars. In terms of the smoothness of the final quasar light curves, the best results are generally obtained by computing these coefficients from a few well exposed stars with a comparable and high signal-to-noise ratio, as opposed to using larger numbers of noisier stars. Furthermore, the ideal normalization stars should be of similar color to the quasar. This minimizes potential differential effects due to varying atmospheric conditions given the relatively broad R band filters used for the monitoring.

In a second step, we rescale the coefficients of each instrument so that a star whose color is closest to the quasar gets the same normalized median flux across all instruments. Residual adjustments to this normalization between instruments will be performed later, using the quasar light curves in temporal regions where the data from different telescopes overlap.

3.3. Photometry of the quasar images

We obtain deblended light curves of the quasar images by *MCS deconvolution photometry*, following Magain et al. (1998). This algorithm fits a single model consisting of (1) some number of point sources and (2) a fine pixel array (hereafter the *pixel channel*) simultaneously to all exposures. In the case of RX J1131–1231, 4 point sources model the quasar images, while the pixel channel represents both the lensing galaxy and the lensed host galaxy (Einstein ring). This model is scaled by the normalization coefficients from Equation 2, and convolved by the exposure-specific PSFs, before being fit to the data. By construction, the relative astrometry of the 4 point sources and the pixel channel are common to all exposures. Only the fluxes of the point sources, the absolute astrometric shift, and a flat residual sky level around the lens are free to vary between exposures. Iteratively, all these parameters are optimized, together with the structure of the regularized pixel channel, so as to minimize a single global χ^2 . This algorithm was already successfully applied in the discovery paper of RX J1131–1231 (Sluse

et al. 2003), as well as in past COSMOGRAIL publications (e.g., Eigenbrod et al. 2007; Vuissoz et al. 2008; Courbin et al. 2011; Sluse et al. 2012) or for similar monitoring data (Burud et al. 2002a; Hjorth et al. 2002; Jakobsson et al. 2005; Morgan et al. 2012). The light curves of the quasar images are a direct output of this procedure.

Errors in the astrometry or the structure of the pixel channel might degrade or bias the photometry of the quasar images. We observe that the influence of the astrometry on the light curves is small. Even for bright quasar images, alterations of the relative position of the point sources by as much as 0''.05 do not significantly modify the light curves. Larger errors tend first to smoothly bias the magnitude measurements, before introducing additional scatter. To obtain light curves for gravitational lenses with image separations on the order of the resolution of the best monitoring data, we find no difference between using the tight astrometric constraints from HST images (Morgan et al. 2006; Chantry et al. 2010) or letting the deconvolution algorithm freely optimize the astrometry of the model.

We reach similar conclusions about the impact of the pixel channel, and its mandatory regularization. Simple numerical experiments show that to first order, the effect of different plausible solutions is similar to very small additive shifts to the flux of the quasar light curves, with only marginally perceptible effects even for faint quasar images. In practice we systematically constrain this pixel channel as well as the point source astrometry using only a subset of images with the best resolution.

3.4. Photometric error estimation

In this section we describe how we compute a rather formal “best case” error estimate for the photometry of each quasar image in each exposure. The normalized flux of a quasar image in a given exposure can be written $f_\star = N_\star \cdot c$, where N_\star is the measured number of photons, and c is the normalization coefficient from Equation 2. We assume that the random error on f_\star has two independent sources: (i) the shot noise σ_{N_\star} of the quasar image itself, and (ii) the noise σ_c of the normalization coefficient as computed in Section 3.2.

We compute σ_{N_\star} following the standard “CCD equation” (see e.g. Howell 2006, chap. 4.4)

$$\sigma_{N_\star} = \sqrt{N_\star + n_{\text{pix}} \cdot (S + R^2)} \quad (3)$$

where S is the sky level, R is the CCD read noise (both in photons per pixel), and n_{pix} is the number of (equivalent) pixels of the software aperture. The dark current is negligible in our images. MCS deconvolution of point sources corresponds to PSF fitting, and following Heyer & Biretta (2004, section 6.5.1) we use for n_{pix} the reciprocal of the PSF *sharpness*,

$$\frac{1}{n_{\text{pix}}} = \sum_{l,m} (\text{PSF}_{lm})^2, \quad (4)$$

where PSF_{lm} is the fraction of light in the total PSF at pixel lm . In this simple “best case” computation, we assume that the PSF is perfectly known, and that the sky level is well determined by the fitting procedure. We do not take into account the surface brightness of the lens and host galaxies, which is generally negligible with respect to the sky level. Furthermore we do not compute noise terms due to the blending with other point sources.

As a sanity check, we have tested our deconvolution photometry algorithm on simulated point source images created with

SkyMaker (Bertin 2009) for various realistic signal-to-noise ratios, elongations, and seeing conditions. Using two or more well-exposed stars to build the PSF, the statistical scatter obtained by our photometry pipeline is only marginally larger than the best-case shot noise computed through the above procedure. Using similar synthetic images, we have also verified that deconvolving two partially blended sources of different fluxes does not significantly bias their measured flux difference in any direction.

To estimate the random uncertainty of the normalization coefficient, we use the standard error of the mean (SEM) of the individual coefficients obtained from the n normalization stars in the exposure (see Equation 1)

$$\sigma_c = \frac{s}{\sqrt{n}} \quad \text{where} \quad s^2 = \frac{1}{n-1} \sum_{i=1}^n (c'_i - \bar{c}')^2. \quad (5)$$

Finally we obtain the error estimated $\sigma_{f_\star \text{ CCD}}$ on the normalized flux through

$$\frac{\sigma_{f_\star \text{ CCD}}^2}{f_\star^2} = \frac{\sigma_{N_\star}^2}{N_\star^2} + \frac{\sigma_c^2}{c^2}. \quad (6)$$

Note that for our COSMOGRAIL and SMARTS monitoring data, the shot noise term from the CCD equation dominates this error budget. The contribution from σ_c is typically $\ll 3\%$ of $\sigma_{f_\star \text{ CCD}}$. Such small calibration errors are important only when measuring very short delays, where they would introduce positively correlated noise into the curves.

3.5. Combination of points per night

In each monitoring night, we observe the lenses m times ($m = 3$ or 5) over ≈ 30 minutes, yielding flux measurements $f_{\star k}$, with $k = 1, \dots, m$ for each quasar image. To reduce the CPU cost and to reject outliers, we bin these measurements by epochs, separately for each instrument and telescope, before measuring the time delays. We attribute to each epoch the *medians* of the photometric measurements $f_{\star k}$ within that night, and the mean of the Heliocentric Julian Dates (HJD).

This approach also allows us to obtain an empirical photometric error estimate for each quasar image, using a measure of the spread of the image's $f_{\star k}$. For increased robustness against outliers, we quantify this spread using the Median Absolute Deviation from the median, or *MAD* (Hoaglin et al. 1983),

$$\text{MAD}(f_{\star k}) = \text{med}(|f_{\star k} - \text{med}(f_{\star k})|). \quad (7)$$

To estimate the usual σ of an (assumed) Gaussian distribution, the MAD is rescaled as

$$\sigma_{f_\star \text{ MAD}} = 1.4826 \cdot \text{MAD}(f_{\star k}). \quad (8)$$

These procedures give us two different error estimates for each epoch and for each quasar image: (1) the median of the errors estimated individually for each exposure $\text{med}(\sigma_{f_\star \text{ CCD}})$, and (2) the more empirical $\sigma_{f_\star \text{ MAD}}$. We observe, as expected, that both error estimations are highly correlated, but also that the empirical $\sigma_{f_\star \text{ MAD}}$ is typically twice as big.

The uncertainty we finally assign to each epoch and quasar image uses the larger of $\text{med}(\sigma_{f_\star \text{ CCD}})$ and $\sigma_{f_\star \text{ MAD}}$. Indeed a realistic error estimate cannot be smaller than either of them. We divide this estimate by the square root of the number of exposures m of the epoch, so that the final error estimate for the median photometric measurement on a given instrument and night is

$$\sigma_{\text{med}(f_\star)} = \frac{\max[\text{med}(\sigma_{f_\star \text{ CCD}}), \sigma_{f_\star \text{ MAD}}]}{\sqrt{m}}. \quad (9)$$

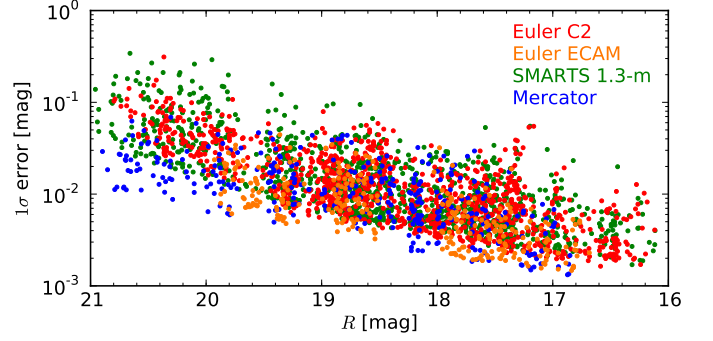


Fig. 3. Photometric 1σ error estimates of each observing epoch as a function of approximate R band magnitude. These errors are estimated following Equation 9.

Fig. 3 displays the distribution of the resulting error estimates as a function of R band magnitude and instrument. The data points from all 4 quasar images of RX J1131–1231 are shown, yielding a broad spread in magnitudes.

3.6. Combination of telescopes

The lens galaxy and the lensed images of the quasar host galaxy differ in color from the quasar itself. As a consequence, the different R filters and CCD response functions of the monitoring instruments might “see” these contaminating sources with slightly different amplitudes even with a perfect calibration of the quasar fluxes. This can result in small mismatches between the light curves obtained using different combinations of telescopes and instruments.

We correct our light curves for such small effects (occasionally as large as 10% of the flux of the faintest quasar images) in this final step. We typically select the instrument with the longest or the highest quality curve as a reference. For RX J1131–1231, this is the SMARTS light curve. For each of the other instruments, we optimize additive magnitude and flux shifts (i.e., multiplicative and additive flux corrections) so to minimize a *dispersion measure* between each instrument's light curve and the reference light curve. We compute this dispersion following the curve shifting technique presented in Tewes et al. (2012), but evaluate it between the light curves of different *instruments*, instead of different quasar images. Provided that the colors of the quasar images are not differentially reddened by absorption in the lens galaxy, we optimize a single *common* magnitude shift per instrument, and individual flux shifts for each quasar image and instrument. This is adequate for RX J1131–1231, as previously suggested by Sluse et al. (2007) and Chartas et al. (2009, 2012).

4. The light curves of RX J1131–1231

All these steps were applied to the COSMOGRAIL and SMARTS data for RX J1131–1231. Fig. 4 shows the final 9-year long light curves of the 4 quasar images. This data is available in machine-readable form at the CDS and on the COSMOGRAIL website². The light curves are dominated by intrinsic quasar variability, with some features on scales as fast as a few weeks. It can be readily seen in the 2008 season for instance, that the delays between A, B and C must be very small, while D is delayed by a bit less than 100 days. Intriguingly, looking only at

² <http://www.cosmograil.org>

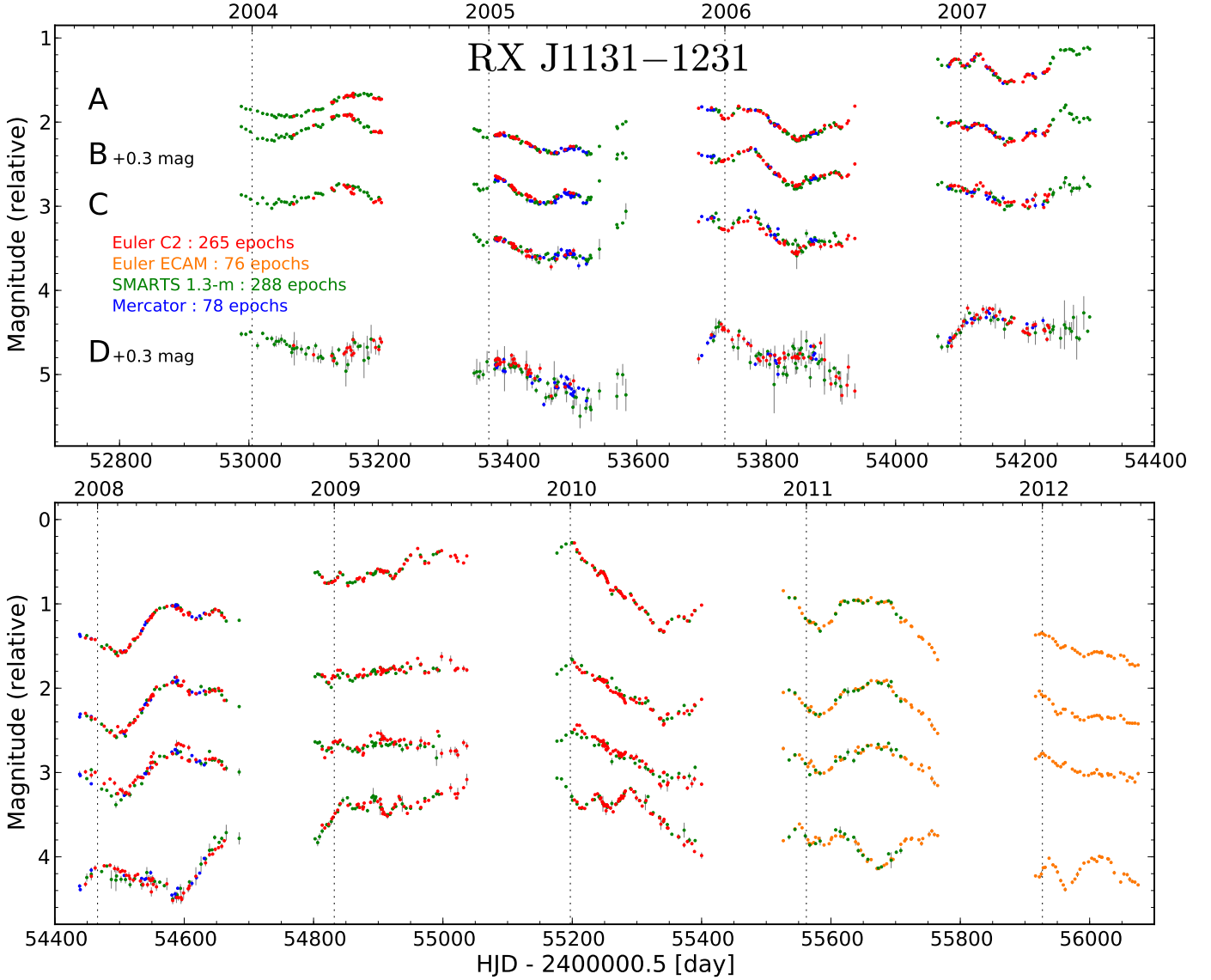


Fig. 4. Optical monitoring of RX J1131–1231, as obtained from deconvolution photometry. From top to bottom are shown the R band light curves for the quasar images A, B, C, D along with the 1σ photometric error bars. Colors encode the contributing instruments. Curves B and D have been shifted by +0.3 magnitudes for display purposes. The light curves are available in tabular form from the CDS and the COSMOGRAIL website.

the first season of A, B and C, one might guess that A is significantly delayed with respect to B and C. We attribute this discrepancy to microlensing variability, which manifestly changes the magnitude difference between the A and B images from the first to the second season. We discuss this “event” further in Section 5.3. Prominent microlensing variability on large time scales is omnipresent, as the flux ratios between the quasar images evolve by as much as a magnitude. These microlensing effects in RX J1131–1231 have been analyzed in Morgan et al. (2010) and Dai et al. (2010).

Lastly, we observe that the photometric error estimates, obtained from equation 9, match the observed scatter well in the smooth sections of curves from the individual telescopes. They are certainly not conservatively large, but we stress that the scale of these error estimates has *no* direct influence on the uncertainties that we will compute for the time delay measurements in the next Section. Our results are robust against a deliberate increase of these error estimates by up to a factor of 5.

5. Time delay estimation

In this section we infer the time delays of RX J1131–1231 from the light curves shown in Fig. 4, closely following the curve shifting and uncertainty evaluation procedures described in Tewes et al. (2012). We summarize the principal ideas below. A major difficulty, and potential source of bias, for curve shifting methods is the presence of “extrinsic” variability in the light curves, on top of the “intrinsic” quasar variability common to all 4 images. The main source of the extrinsic signal is variable microlensing magnification due to the motions of the stars in the lens galaxy. As shown by Mosquera & Kochanek (2011), microlensing can affect light curves over a broad range of time scales. For RX J1131–1231, Mosquera & Kochanek (2011) estimate a time scale of ≈ 11 years for the crossing of a stellar Einstein radius, and ≈ 3 months for the source radius to cross a caustic.

The curve shifting methods of Tewes et al. (2012) try to minimize the bias due to such extrinsic variability in different ways.

They all rely on iterative optimizations of time shifts of the 4 light curves, and yield self-consistent point estimates of the time delays between all 6 image pairs :

1. **The free knot spline technique** simultaneously fits one common *intrinsic* spline for the quasar variability, and independent, smoother *extrinsic* splines for the microlensing to the light curves. The curves are shifted in time to optimize this fit. This method is similar to the polynomial method of Kochanek et al. (2006).
2. **The regression difference technique** shifts continuous regressions of the curves in time, to minimize the *variability* of the differences between them. This novel method does not involve an explicit model for the microlensing variability.
3. **The dispersion-like technique**, inspired by Pelt et al. (1996), shifts the curves so as to minimize a measure of the *dispersion* between the overlapping data points. This method has no explicit model for the common intrinsic variability of the quasar, but it includes polynomial models for the extrinsic variability.

Using several independent algorithms allows us to cross-check our estimates for technique-dependent biases. Indeed, as important as the time delay point estimates themselves is the reliable estimation of their uncertainties. For this we follow a Monte Carlo approach, by applying the curve shifting techniques to a large number of synthetic light curves with *known* time delays. These curves are drawn from a light curve model that mimics both the observed intrinsic and extrinsic variability of the real observations, randomizing only the unrecoverable short time scale extrinsic variability. This latter “correlated noise” locally adapts its amplitude to the scatter of the observed data points. As a result, the synthetic curves are virtually indistinguishable from the real ones (see Figs. 5 and 6 of Tewes et al. 2012, for an illustration).

5.1. Application to RX J1131–1231

We begin by evaluating the robustness of the time shift optimization of each method given the data. For this we repeatedly run the delay point estimators on the observed light curves. Each time, we start the optimizations from random initial conditions, using time shifts uniformly drawn ± 10 days around plausible solutions. The resulting distributions of point estimates are shown in Fig. 5, characterizing what we call the *intrinsic variance* of the estimates.

These tests are essential to check the reliability of the non-linear optimization algorithms for a given data set and model parameters. We stress that this procedure should not be interpreted as importance sampling of any posteriors. Generally speaking, overly flexible models dilute the time delay information and yield higher intrinsic variances. For the free knot spline technique we have chosen an average knot step of 20 days for the quasar variability spline, and 150 days for the extrinsic splines. For the dispersion-like method we model the extrinsic variability by independent linear trends on each season. If the light curves sufficiently constrain the time delays, the free knot spline and regression difference techniques can easily be adjusted to display small intrinsic variance and roughly unimodal distributions, as in Fig. 5. The dispersion-like technique is inherently more sensitive to initial conditions, due to a much higher roughness of the scalar objective function that it minimizes. As discussed in Tewes et al. (2012), it is important to note that smoothing the objective function does not necessarily lead to more accurate time delay estimates. For each technique, we use the mean values of

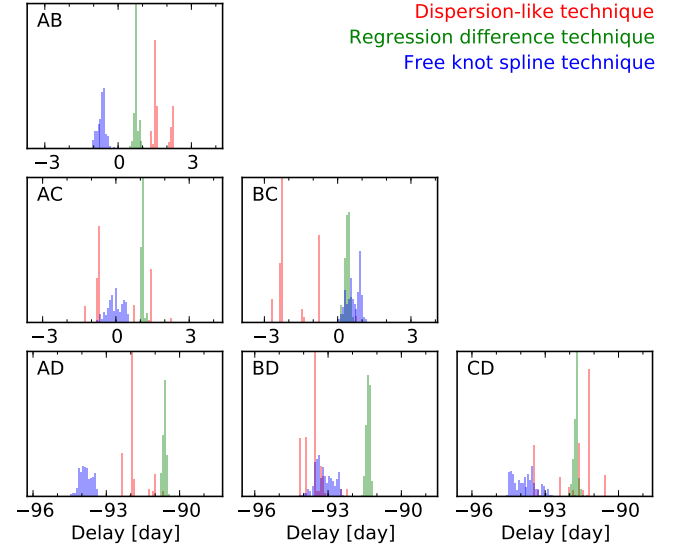


Fig. 5. Analysis of the “intrinsic variance” of the time delay point estimators, found by running the three curve shifting techniques 200 times on the light curves shown in Fig. 4, starting the optimizations from random initial time shifts. The *widths* of these distributions reflect the failure of the methods to converge to a single optimal solution given the data. We stress that these distributions do *not* represent probability density functions the time delays, and that a small intrinsic variance does *not* imply that a time delay estimation is precise or accurate, only that the error surface is relatively smooth.

the distributions in Fig. 5 as our best time delay estimates. They correspond to the points in Fig. 6, which summarizes our results for RX J1131–1231.

The remaining part of the analysis is solely about estimating realistic error bars for these point estimates. For this we proceed by blindly running the exact same curve shifting techniques on 1000 sets of fully synthetic light curves with true time delays randomly chosen around the measured ones. Again, we start the methods from random initial shifts. Statistics of the resulting time delay measurement errors (i.e., measurement – truth), are shown in Fig. 7 as a function of the true delays of the synthetic curves. In this figure, the shaded rods show the mean measurement error, which we call systematic error or bias, while the error bars indicate the standard deviation of the measurement errors, representing the random errors. The total errors for our delay measurements, shown by the error bars in Fig. 6, are computed by adding the maximum bias and the maximum random error of each panel in quadrature. Note that we run the curve shifting techniques only once on each random synthetic curve set, so this error analysis takes into account the intrinsic variance. This is rather conservative, given that we use the mean result of several time shift optimizations as our best estimates for the observed data. However, the additional contribution to the uncertainty estimates is negligible for methods with low intrinsic variance. Before discussing these results, we show in Fig. 8 the same measurement errors as in Fig. 7, but plotted for each delay against each other, in order to explore potential abnormal correlations.

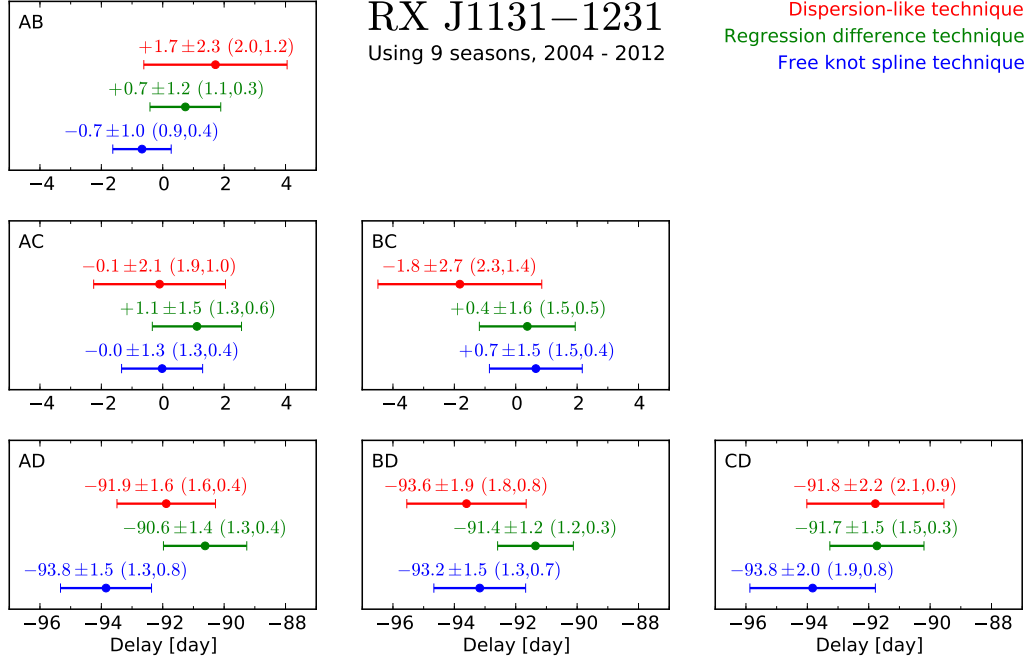


Fig. 6. Time delay measurements along with the 1σ errors for RX J1131–1231 obtained by our 3 standard techniques from the full 9-season long light curves shown in Fig. 4. The random and systematic error contributions are given in parentheses for each delay. The error bars represent the random plus systematic errors summed in quadrature. A positive AB-delay Δt_{AB} means that image B leads image A. We consider the measurements from the *regression difference technique*, which display the smallest bias and variance in our error analysis, as the delays that should be used to constrain cosmology and lens models.

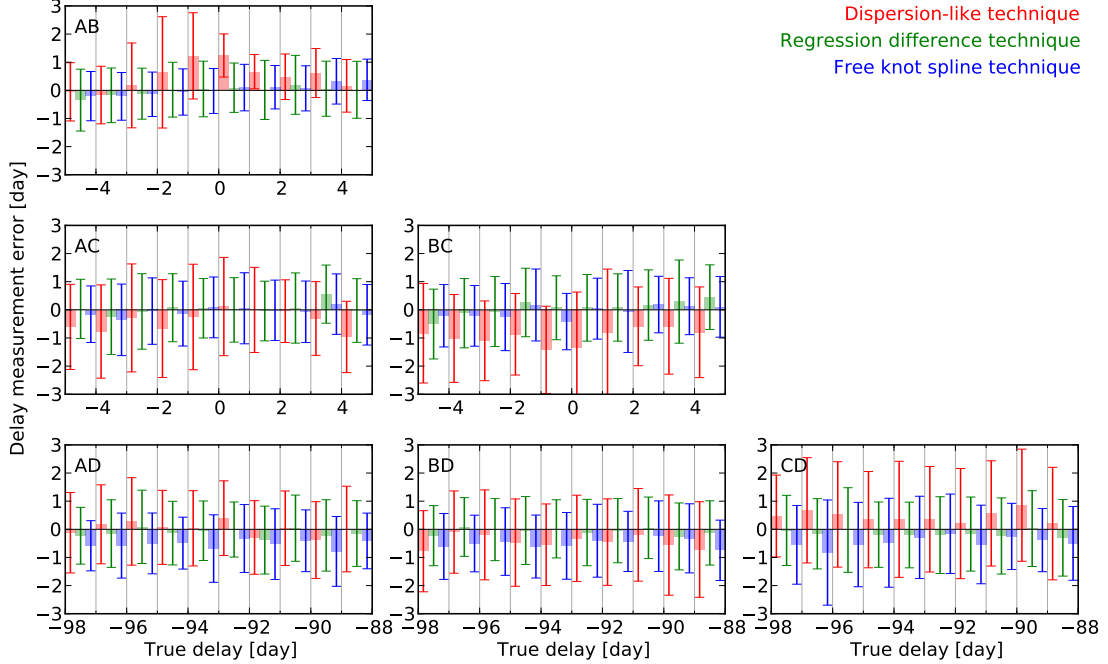


Fig. 7. Results of the Monte Carlo analysis leading to the error estimates for our time delay measurements shown in Fig. 6. We obtain our uncertainty estimates by applying the curve shifting techniques to 1000 synthetic light curve sets that closely mimic the observed data but have known “true” time delays. The vertical axes show the delay measurement error, as compared to the true delays used to generate the synthetic curves (horizontal axes). Separately for each panel, the outcomes are binned according to the true time delays. The bin intervals are shown as light vertical lines. Within each bin, the shaded rods and error bars show the systematic and random errors, respectively, of the delay measurements for each technique.

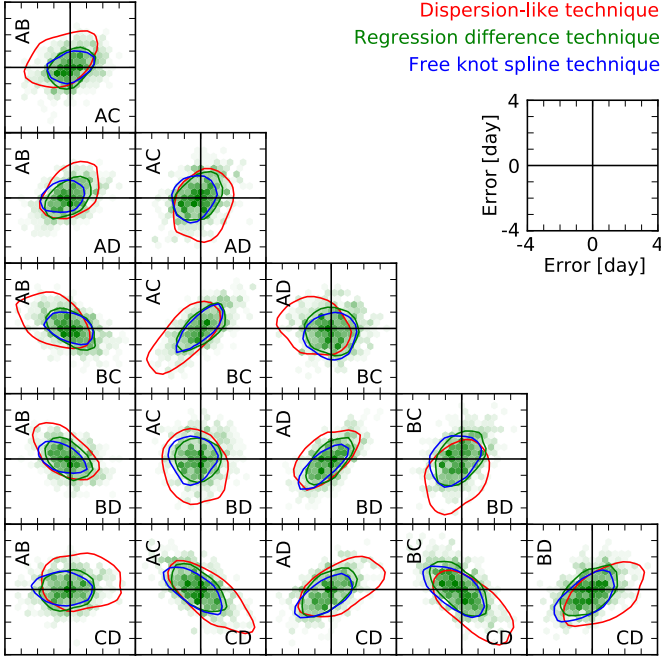


Fig. 8. Correlations of delay measurement errors for synthetic light curves mimicking RX J1131–1231. The measurement errors are the same as shown in Fig. 7, but this time marginalizing over the true delays. Crosshairs indicate zero error, and the inset shows the scale of each panel. For clarity, only single contours at half of the maximum density are shown for two of the techniques. We observe no correlations (i.e., oblique contours) between the unrelated delay measurements along the short diagonal of this figure.

5.2. Discussion

In terms of simple lens model considerations, RX J1131–1231 is a “long axis cusp lens”, with the source is located inside a cusp of the tangential caustic curve on the long axis of the potential (Sluse et al. 2003). Images A and D are the saddle points of the arrival time surface, while B and C are minima (Blandford & Narayan 1986). Fig. 14 of Saha et al. (2006) gives an illustration of this particular lens. Hence the delays Δt_{AB} and Δt_{AC} are predicted to be small but positive, while Δt_{AD} is negative and large; the possible arrival time orders are BCAD or CBAD.

Our measured delays, as shown in Fig. 6, are consistent with these predictions, although the delays between A, B and C are compatible with being zero given the 1σ errors. It is reassuring to observe that the three methods yield consistent results for all delays despite the very different methodologies, which we see as a success of our error estimation procedure. Keep in mind however that the delay measurements are not independent, because they all use the same single set of observed light curves.

Let us look in more detail at the measurement error analysis shown in Fig. 7, which is based solely on synthetic curves. Overall, we do not observe any strong dependence of a method’s bias or random error on the true time delays used in the simulations. For most image pairs and methods, the random errors are more important than the almost negligible bias. Exceptions are the two results from the dispersion-like technique, which is observed to overpredict Δt_{AB} , and underpredict Δt_{BC} , by about one day. Remarkably, this same technique also measured the highest (lowest) delay Δt_{AB} (Δt_{BC}) for the observed curves (Fig. 6), when compared to the other methods. We see this as a further in-

dication that the bias estimates are reliable. Finally, Fig. 8 shows no sign of abnormal correlations between measurement errors of quasar image pairs, whether they share a common quasar image or not. By construction our time delay uncertainty estimates for each image pair marginalize over all other pairs.

Which delay measurement method performs best on RX J1131–1231? Given that the regression difference technique yields both the smallest biases and the smallest random errors, we simply conclude that its measurements are the most informative ones. We will use the delays from the regression difference technique, expressed with respect to quasar image B, to measure the time-delay distance towards RX J1131–1231. Details and results of lens modeling, as applied to RX J1131–1231, are described in detail in Suyu et al. (2012a).

5.3. Delay measurement on subsets of the light curves

Our long light curves, featuring several delay-constraining intrinsic variability patterns, suggest that we should independently measure the time delays from subsets of the observing seasons. This analysis represents an invaluable check of the consistency of the time delay estimation procedure of Tewes et al. (2012), and hence for the results from the full light curves.

We analyze 3 subsets of the available data : (1) the first two seasons, (2) the first 5 seasons, and (3) the remaining last 4 seasons. Case (1) is chosen for comparison with Morgan et al. (2006). We perform the analyses *from scratch*, without using any knowledge about the delay estimates from the full curves. In particular, for each case, we build a new model for the synthetic light curves, independently adjusted so that they best resemble the observations in that time period based on the statistical criteria presented in Tewes et al. (2012, Section 7). We do not alter any parameters of the curve-shifting techniques.

Fig. 9 presents the resulting delay measurements. The data points and error bars depict the individual point estimates and the corresponding 1σ total errors obtained by each technique for the 3 cases. The shaded regions show the 1σ intervals from the full 9 seasons taken from Fig. 6. We observe :

1. Using only the first 2 seasons, our three methods are systematically biased towards large values of Δt_{AB} and Δt_{AC} . Our interpretation is that some microlensing variability in image A conspiratorially imitates a time shift, particularly around mid 2004. The estimates from the dispersion-like and spline technique roughly reproduce the results obtained from the same two seasons by Morgan et al. (2006) $\Delta t_{AB} = 12.0 \pm 1.5$, $\Delta t_{AB} = 9.6 \pm 1.9$ and $\Delta t_{AD} = -87 \pm 8$, but with significantly wider, yet still too small, error bars. This demonstrates the need of long monitoring programs to measure accurate time delays and minimize the influence of unfortunately placed microlensing events.
2. For these same two seasons, the regression difference method yields far better estimates for these delays – including adequately sized error bars that encompass the delays estimated from the full light curves. Note that the regression difference technique is the only one that does not involve an explicit *model* for the microlensing.
3. For the other 2 divisions of the data, namely seasons 1-5 and 6-9, which are disjoint and hence independent, our methods yield consistent results. As would be expected, both subcases show larger error bars than the combined analysis of all seasons.

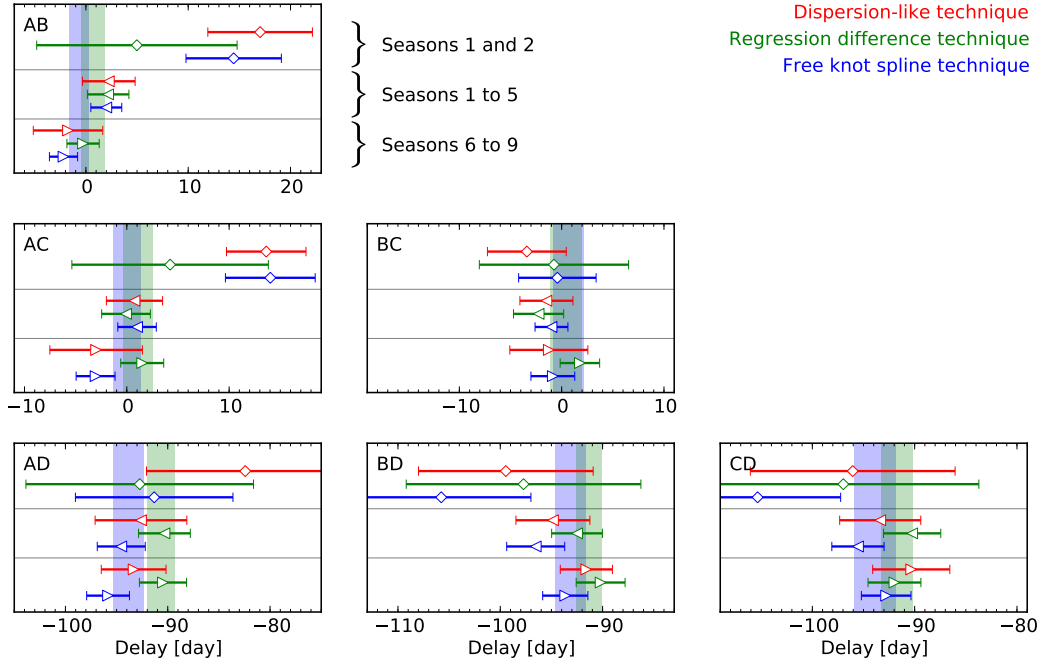


Fig. 9. Application of the curve shifting techniques to subsections of the full light curves. The square diamonds (top) show measurements using only the 2 first seasons, the leftward triangles (middle) use the first 5 seasons, and the rightward triangles (bottom) use the last 4 seasons. All error bars depict *total* errors, as in Fig. 6. The blue and green shaded regions correspond to the interval covered by the total error bars using the full 9 seasons, by the spline and regression difference techniques, respectively.

6. Conclusions

The first part of this paper describes the COSMOGRAIL data reduction procedure, which will be used to reduce all the data gathered by our monitoring campaign of gravitationally lensed quasars. In the second part we apply this pipeline to our COSMOGRAIL and SMARTS observations of the quad lens RX J1131–1231, leading to an unprecedented set of 9 year long light curves of high photometric quality. Several strong and fast intrinsic quasar variability patterns constrain the time delays between the multiple images. Microlensing-related “extrinsic” variability is clearly present, as pointed out and analyzed in previous studies (Sluse et al. 2006, 2007; Morgan et al. 2010; Dai et al. 2010; Chartas et al. 2012). However, this distorting signal does not prevent us from measuring accurate time delays, using the three independent algorithms of Tewes et al. (2012).

The best time delay estimates of RX J1131–1231 are provided by the regression difference technique. It measures the 91-day delays between D and the other quasar images to a fractional 1σ uncertainty of 1.5%. This error estimate is obtained by applying the techniques to synthetic curves with known time delays, which contain extrinsic variability features similar to the observed ones. We demonstrate the consistency of our error estimates by independently measuring time delays – including error bars – from subsets of the observed light curves of RX J1131–1231. This experiment also reveals that long multi-year monitoring is essential to reliably measure time delays, despite progress on the methods.

The results from this paper are used to constrain the time delay distance towards RX J1131–1231, and deduce stringent implications for cosmology in Suyu et al. (2012a). With our time delay measurement errors of only 1.5%, the accuracy of this cosmological inference is actually limited by the residual uncertainty of (1) the gravitational potential of the lens galaxy

and (2) the large scale structures along the line of sight to the quasar (Bar-Kana 1996).

Acknowledgements. We are deeply grateful to the numerous observers who contributed to the data acquisition at the Swiss Euler telescope, the SMARTS 1.3-m telescope, and the Flemish-Belgian Mercator telescope. We wish to thank Sherry Suyu and Tommaso Treu for precious discussions and comments. COSMOGRAIL is financially supported by the Swiss National Science Foundation (SNSF). We also acknowledge support from the International Space Science Institute in Bern, where some of this research has been discussed. CSK and AMM are supported by NSF grant AST-1009756. EE was partially supported by ESA and the Belgian Federal Science Policy (BELSPO) in the framework of the PRODEX Experiment Arrangement C-90312. DS acknowledges support from the Deutsche Forschungsgemeinschaft, reference SL172/1-1.

References

- Bar-Kana, R. 1996, *ApJ*, 468, 17
- Bertin, E. 2009, *Memorie della Società Astronomica Italiana*, 80, 422
- Bertin, E. & Arnouts, S. 2010, *Astrophysics Source Code Library*, record ascl:1010.064
- Blackburne, J. A., Kochanek, C. S., Chen, B., Dai, X., & Chartas, G. 2011, *arXiv*, 1112.0027
- Blandford, R. & Narayan, R. 1986, *ApJ*, 310, 568
- Burud, I., Courbin, F., Magain, P., et al. 2002a, *A&A*, 383, 71
- Burud, I., Hjorth, J., Courbin, F., et al. 2002b, *A&A*, 391, 481
- Chang, K. & Refsdal, S. 1979, *Nature*, 282, 561
- Chantry, V., Sluse, D., & Magain, P. 2010, *A&A*, 522, 95
- Chartas, G., Kochanek, C. S., Dai, X., et al. 2012, *arXiv*, 1204.4480
- Chartas, G., Kochanek, C. S., Dai, X., Poindexter, S., & Garmire, G. 2009, *ApJ*, 693, 174
- Courbin, F., Chantry, V., Revaz, Y., et al. 2011, *A&A*, 536, A53
- Dai, X., Kochanek, C. S., Chartas, G., et al. 2010, *ApJ*, 709, 278
- DePoy, D. L., Atwood, B., Belville, S. R., et al. 2003, in *Proceedings of SPIE (SPIE)*, 827–838
- Eigenbrod, A., Courbin, F., & Meylan, G. 2007, *A&A*, 465, 51
- Fassnacht, C. D., Xanthopoulos, E., Koopmans, L., & Rusin, D. 2002, *ApJ*, 581, 823
- Fohlmeister, J., Kochanek, C. S., Falco, E. E., Morgan, C. W., & Wambsganss, J. 2008, *ApJ*, 676, 761

- Fohlmeister, J., Kochanek, C. S., Falco, E. E., et al. 2012, arXiv, 1207.5776
- Heyer & Biretta. 2004, WFPC2 Instrument Handbook, STScI, Baltimore
- Hjorth, J., Burud, I., Jaunsen, A. O., et al. 2002, ApJ, 572, L11
- Hoaglin, D. C., Mosteller, F., & Tukey, J. W., eds. 1983, Understanding Robust and Exploratory Data Analysis (Wiley)
- Howell, S. B. 2006, Handbook of CCD Astronomy, 2nd edn. (Cambridge University Press)
- Jakobsson, P., Hjorth, J., Burud, I., et al. 2005, A&A, 431, 103
- Kochanek, C. S. 2002, ApJ, 578, 25
- Kochanek, C. S., Morgan, N. D., Falco, E. E., et al. 2006, ApJ, 640, 47
- Kundic, T., Turner, E. L., Colley, W. N., et al. 1997, ApJ, 482, 75
- Linder, E. V. 2011, Physical Review D, 84, 123529
- MacLeod, C. L., Ivezić, Z., Kochanek, C. S., et al. 2010, ApJ, 721, 1014
- Magain, P., Courbin, F., & Sohy, S. 1998, ApJ, 494, 472
- Morgan, C. W., Eyler, M. E., Kochanek, C. S., et al. 2008a, ApJ, 676, 80
- Morgan, C. W., Hainline, L. J., Chen, B., et al. 2012, ApJ, 756, 52
- Morgan, C. W., Kochanek, C. S., Dai, X., Morgan, N. D., & Falco, E. E. 2008b, ApJ, 689, 755
- Morgan, C. W., Kochanek, C. S., Morgan, N. D., & Falco, E. E. 2010, ApJ, 712, 1129
- Morgan, N. D., Kochanek, C. S., Falco, E. E., & Dai, X. 2006, arXiv, astro-ph/0605321
- Mosquera, A. M. & Kochanek, C. S. 2011, ApJ, 738, 96
- Moustakas, L. A., Abazajian, K., Benson, A., et al. 2009, Astro2010: The Astronomy and Astrophysics Decadal Survey Science White Papers
- Pelt, J., Kayser, R., Refsdal, S., & Schramm, T. 1996, A&A, 305, 97
- Press, W. H., Rybicki, G. B., & Hewitt, J. N. 1992, ApJ, 385, 404
- Refsdal, S. 1964, MNRAS, 128, 307
- Saha, P., Courbin, F., Sluse, D., Dye, S., & Meylan, G. 2006, A&A, 450, 461
- Schild, R. & Thomson, D. J. 1997, Astronomical Journal, 113, 130
- Schmidt, R. W. & Wambsganss, J. 2010, General Relativity and Gravitation, 42, 2127
- Sluse, D., Chantry, V., Magain, P., Courbin, F., & Meylan, G. 2012, A&A, 538, A99
- Sluse, D., Claeskens, J.-F., Altieri, B., et al. 2006, A&A, 449, 539
- Sluse, D., Claeskens, J.-F., Hutsemekers, D., & Surdej, J. 2007, A&A, 468, 885
- Sluse, D., Surdej, J., Claeskens, J.-F., et al. 2003, A&A, 406, L43
- Suyu, S. H., Auger, M. W., Hilbert, S., et al. 2012a, ApJ, submitted
- Suyu, S. H., Marshall, P. J., Auger, M. W., et al. 2010, ApJ, 711, 201
- Suyu, S. H., Marshall, P. J., Blandford, R. D., et al. 2009, ApJ, 691, 277
- Suyu, S. H., Treu, T., Blandford, R. D., et al. 2012b, arXiv, 1202.4459
- Tewes, M., Courbin, F., & Meylan, G. 2012, arXiv, 1208.5598
- van Dokkum, P. G. 2001, PASP, 113, 1420
- Vanderriest, C., Schneider, J., Herpe, G., et al. 1989, A&A, 215, 1
- Vuissoz, C., Courbin, F., Sluse, D., et al. 2008, A&A, 488, 481
- Vuissoz, C., Courbin, F., Sluse, D., et al. 2007, A&A, 464, 845
- Wyrzykowski, L., Udalski, A., Schechter, P. L., et al. 2003, Acta Astronomica, 53, 229